

Universidad Complutense de Madrid
Facultad de CC. Químicas

***Ajustes de modelos y operaciones con
matrices en EXCEL***

Luis Vicente Pérez Arribas
Dpto. de Química Analítica
Fac. de CC. Químicas

Madrid 2018

Convenciones tipográficas utilizadas en este tutorial

Dentro del texto aparecerán palabras escritas en una tipografía diferente del conjunto del texto. Esta tipografía se ha utilizado básicamente para resaltar algunos elementos de cierta importancia. Los tipos de letra utilizados han sido los siguientes:

Cursiva. Se utilizará para términos anglosajones o de uso poco común.

Negrita. En este estilo se mostrarán nombres de botones y combinaciones de teclas para realizar determinadas acciones. También se utiliza para resaltar el texto o nombres de programas informáticos.

Courier. Para mostrar nombres de funciones, referencias a celdillas y códigos y fórmulas.

Arial. Para destacar los elementos relativos a la interfaz de usuario (nombre de una opción de menú, título de una ventana, etc.).

Operaciones básicas con la hoja de cálculo

Microsoft Excel, como todas las hojas electrónicas de cálculo, permite una gran variedad de operaciones aritméticas, desde las más sencillas hasta operaciones que requieren el empleo de complicadas fórmulas y ecuaciones matemáticas. En el caso de las operaciones aritméticas sencillas, Excel trabaja de forma similar a una calculadora, en la que se introducen los datos numéricos y los operadores aritméticos para finalmente obtener el resultado tras pulsar **Intro**, [=]. Supóngase que se quisiera calcular la masa molecular del ácido nítrico, HNO₃. Puesto que la masa atómica del nitrógeno es 14; la del oxígeno, 16 y la del hidrógeno, 1, lo único que habría que hacer es escribir en la Barra de fórmulas:

$$=1+14+16*3$$

Tras pulsar **Intro**, [=] aparecerá en la celda activa el resultado de la operación, 63. El símbolo [=] sirve para indicar a Excel que lo que va a continuación es una fórmula matemática y que sólo tiene que presentar el resultado de las operaciones indicadas. De no hacerse así, Excel considera que se trata de la introducción de una serie de datos alfanuméricos y los presentará como tales, sin realizar operación matemática alguna. En la siguiente tabla se muestran los operadores aritméticos que se utilizan para cada tipo de operación aritmética:

Tabla 1. *Operadores aritméticos*

Operador	Operación	Ejemplo	Resultado
+ (signo más)	Suma	=3+3	6
- (signo menos)	Resta (también valor negativo)	=-3-1	-4
* (asterisco)	Multiplicación	=3*5	15
/ (barra oblicua)	División	=1/3	0,3333333
^ (acento circunflejo)	Potencia	=5^2	25
% (signo de porcentaje)	Porcentaje	=20%	0,20

Las operaciones matemáticas realizadas con la hoja de cálculo no tienen por qué ser tan simples como las del ejemplo de la masa molecular del ácido nítrico. Hay casos en que en una misma expresión matemática pueden aparecer múltiples operandos con sus respectivos operadores. En estos casos hay que tener en cuenta el orden de prioridad en que se realizarán las diferentes operaciones, ya que, de lo contrario puede verse afectado el resultado. Excel sigue las reglas estándar de preferencias de los operadores, es decir, las operaciones exponenciales se realizan primero, la multiplicación y la división después, con igual prioridad y finalmente, la adición y la sustracción, también de igual prioridad. Si en la fórmula introducida en la Barra de fórmulas se encuentra un signo menos (-), detrás de otro operador o sin que exista otro número a su izquierda, Excel lo considerará como un número negativo (p.e. $=5*-2$ retorna como resultado -10). Si una fórmula contiene operadores con la misma prioridad (por ejemplo, si una fórmula contiene un operador de multiplicación y otro de división), Excel evaluará primero el operador que se encuentre más a la izquierda y continuará hacia la derecha (p.e. $=10/2*5$ dará como resultado 25 y $=10*2/5$ dará 4).

Este orden puede modificarse utilizando paréntesis, pues las operaciones contenidas entre paréntesis se efectúan primero. Cuando dentro de un paréntesis se encuentran varios operadores aritméticos, estos se ejecutan según las reglas de prioridad comentadas anteriormente. El siguiente caso es un ejemplo de operación matemática que requiere el empleo de paréntesis. Supóngase que se quiere calcular las raíces de una ecuación de segundo grado. Como es sabido, las raíces de las ecuaciones de segundo grado, $ax^2 + bx + c = 0$, se calculan mediante la expresión matemática:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Si se quisiera calcular las raíces de la ecuación $x^2 + 2x - 3 = 0$, en la barra de fórmulas habría que introducir la expresión

$$=(-2+(2^2-4*1*-3)^(1/2))/2$$

para una de las raíces y

$$=(-2-(2^2-4*1*-3)^(1/2))/2$$

para la otra. Si se han tenido en cuenta las prioridades de los operadores y se han utilizado correctamente los paréntesis, los resultados que se obtendrán serán 1 y -3, respectivamente.

Además de los operadores aritméticos reseñados en la tabla 1, Excel dispone de otros operadores, muy útiles en otros tipos de operaciones. Son los operadores de comparación, que permiten establecer la relación entre dos operandos (Tabla 2).

Tabla 2. Operadores de comparación

Operador de comparación	Significado	Ejemplo
= (signo igual)	Igual a	A1=B1
> (signo mayor que)	Mayor que	A1>B1
< (signo menor que)	Menor que	A1<B1
>= (signo mayor o igual que)	Mayor o igual que	A1>=B1
<= (signo menor o igual que)	Menor o igual que	A1<=B1
<> (signo distinto de)	Distinto de	A1<>B1

Estos operadores se emplean principalmente para comparar valores contenidos en diferentes celdas y generan un valor lógico: VERDADERO o FALSO, aunque también pueden ser utilizados en los cálculos, ya que VERDADERO equivale a 1 y FALSO a 0 (p.e. la fórmula $= (B1=B2) * 5$ daría 0 como resultado si el valor contenido en la celda B1 es diferente al contenido en la B2, o 5 si son iguales).

Empleo de funciones y fórmulas predefinidas

Excel contiene una amplia colección de funciones predefinidas que permiten en muchos casos hacer cálculos de forma sencilla sin necesidad de teclear la ecuación correspondiente. Existen funciones para casi todo tipo de operaciones rutinarias, desde simples sumas hasta complejos cálculos financieros o de ingeniería. Las hay matemáticas, trigonométrica, estadísticas, financieras, etc. Para conocer las funciones disponibles en Excel, hay que pulsar el botón **Insertar función**, que se encuentra en la Barra de fórmulas marcado con el icono **[fx]**. Una vez pulsado, aparecerá una ventana emergente con una lista desplegable que mostrará las diferentes funciones agrupadas por categorías.

También se puede acceder a las diferentes funciones disponibles en Excel abriendo la ficha Fórmulas en la Cinta de opciones. Dentro del grupo Biblioteca de funciones se encuentra destacado el icono **[fx]** o pulsando con el cursor del ratón sobre alguna de las categorías que se muestran junto a dicho icono.

Algunas funciones útiles para el cálculo numérico son: SENO, COS, TAN; PI; LN, LOG10; EXP y RAIZ. Estas y otras funciones no solo sirven para cálculos numéricos sencillos, sino que pueden formar parte de una fórmula matemática, lo que simplifica en gran medida su edición en la Barra de fórmulas. Así, en el ejemplo anterior del cálculo de las raíces de las ecuaciones de segundo grado, podría haberse combinado los operadores matemáticos con la función RAIZ¹, de forma que las fórmulas habrían quedado:

$$= (-2 + \text{RAIZ} (2^2 - 4 * 1 * (-3))) / 2$$

y

$$= (-2 - \text{RAIZ} (2^2 - 4 * 1 * (-3))) / 2$$

La mayoría de las funciones existentes necesitan argumentos sobre los que operar. Los argumentos son elementos que proporcionan información a la función. Este argumento puede ser un valor numérico, el valor contenido en una celdilla, los valores de un grupo o rango de celdillas u otra fórmula matemática. En el caso del ejemplo anterior, el argumento utilizado con la función RAIZ es $2^2 - 4 * 1 * (-3)$. Algunas funciones requieren más de un argumento. En estos casos, se introducen dentro de los paréntesis, separados por punto y coma y de acuerdo a un orden preestablecido. También existen funciones que no requieren argumento o que su uso es opcional. En estos casos también es necesario escribir los paréntesis, aunque sin colocar nada en su interior (p.e. para operar con el número π , la función será $=\text{PI}()$). En caso de no utilizar los paréntesis, Excel devolverá el mensaje de error #¿NOMBRE?).

¹ Cuando se utiliza la función RAIZ con un argumento que es una fórmula, hay que tener cuidado de que esta no genere un número negativo cuya raíz cuadrada sería un número imaginario. Aunque Excel dispone de funciones para trabajar con números imaginarios y complejos, como IMAGINARIO, IM.SUM, IM.REAL, etc. no reconoce estos números cuando se generan en un cálculo, devolviendo el mensaje de error #¡NUM!

Ajuste de series de datos a modelos matemáticos sencillos

Además de representar series de datos, en química (y en otras ciencias) es muy frecuente ajustar dichos datos a modelos o ecuaciones matemáticas más o menos complejas. Estos modelos matemáticos se conocen como modelos de regresión y su objetivo es estudiar y conocer la relación cuantitativa entre una variable dependiente, casi siempre una respuesta o resultado experimental, y una o varias variables independientes, cuyos valores suelen ser fijados por el experimentador. En el trabajo experimental en química son frecuentes los casos en que se toman medidas a intervalos de tiempo regulares, a temperaturas previamente fijadas o a presiones preestablecidas. El tiempo, la temperatura o la presión serían las variables independientes y las medidas deducidas de las lecturas de los aparatos o calculadas por el experimentador, serían las variables dependientes, pues su valor es función, en cada caso, de los valores previos fijados para las variables independientes. A continuación, se mostrará en unos sencillos pasos cómo realizar estos ajustes. Supóngase que se ha obtenido en el laboratorio los datos de la tabla 3 y que se desean ajustar a algún modelo matemático.

Tabla 3. Ejemplo de datos para ajustar a modelo matemático

x	y
0,352	1,09
0,803	1,78
1,03	2,6
1,38	3,03
1,75	4,01

Como paso previo, antes de proceder al ajuste de los datos, es muy importante representarlos en un gráfico, ya que su representación puede dar una orientación sobre el modelo matemático que podría ser más adecuado. Para estos casos, el tipo de gráfico más adecuado es el denominado XY (Dispersión), que compara los pares de valores y los representa a escala. Se puede acceder a este tipo de representación gráfica a través de la galería de tipos de gráficos o presionando directamente sobre el botón Dispersión, en el grupo Gráficos de la Cinta de opciones. Dentro de los diferentes subtipos que aparecen en la ventana, se escogerá el primero, denominado Dispersión solo con marcadores, que es el correspondiente a los puntos sin unir por líneas (figura 1). Una vez seleccionados los datos que se van a representar y seguir los pasos indicados anteriormente, aparecerá sobre la hoja de cálculo la gráfica con los puntos, lo que permitirá constatar que los datos se ajustan bastante bien a una línea recta. El siguiente paso será obtener los coeficientes de ajuste al modelo matemático deseado, en este caso el modelo polinómico de primer orden, cuya ecuación es la línea recta:

$$y = a \cdot x + b$$

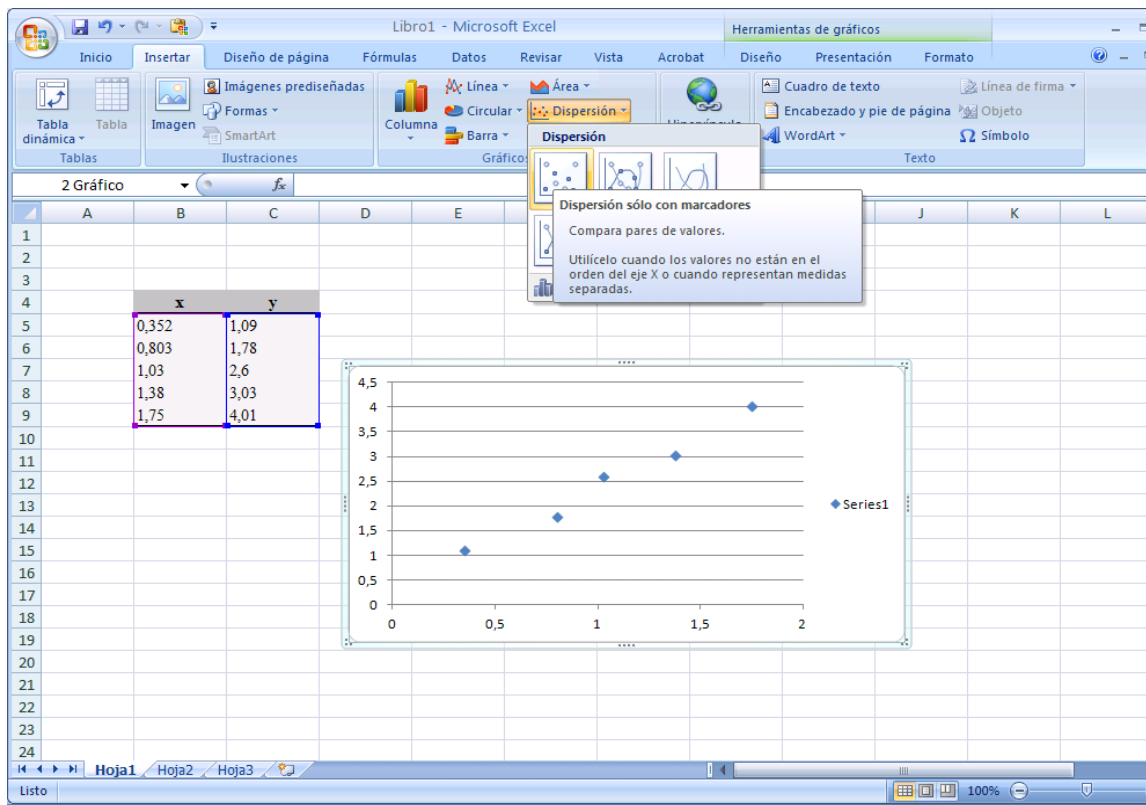


Figura 1. Ejemplo de representación gráfica de datos para el tipo gráfico XY(Dispersión).

Obtención de los coeficientes mediante el empleo de funciones estadísticas de Excel

El polinomio de primer orden o línea recta está definido por la pendiente (coeficiente a en la ecuación anterior) y el valor de la ordenada en el origen (coeficiente b). Además de estos dos coeficientes, es frecuente calcular el denominado **coeficiente de regresión**, r , que es un parámetro de estimación de la calidad del ajuste de los datos experimentales al modelo matemático propuesto. Cuanto más se aproxime su valor a 1, (o a -1 si la pendiente es negativa) más fiable es el modelo, pues en este caso la correlación sería perfecta, con todos los puntos experimentales situados sobre la recta calculada. Sin embargo, cada vez es más frecuente el empleo del **coeficiente de determinación** o **coeficiente de correlación múltiple**, R^2 , que sirve para cualquier modelo de ajuste, no solo para el lineal, y que en el caso de la línea recta es igual a r^2 . Desde el punto de vista estadístico, R^2 indica el porcentaje de ajuste entre modelo matemático y los datos experimentales que lo han generado, de forma que un valor del coeficiente de correlación superior a 0,99 implica que el modelo se ajusta en más de un 99% a los datos. Estos coeficientes se pueden determinar de forma muy sencilla, ya que Excel dispone de varias funciones que dan directamente los valores de la pendiente, la ordenada y el coeficiente correlación. Estas funciones son: PENDIENTE, INTERSECCION.EJE, COEF.DE.CORREL y COEFICIENTE.R2. Para utilizarlas, se introducen en una hoja Excel los datos de x y y . Seguidamente se selecciona la celda donde se quiere mostrar la pendiente de la recta. Una vez seleccionada la celda, se pulsa sobre el botón **Insertar función**, que se encuentra en la Barra de fórmulas marcado con el icono $[fx]$. Una vez abierta la ventana de dialogo, se selecciona la categoría Estadísticas o Todas y se

busca `PENDIENTE` en la lista. Se introduce el rango de celdas en **Conocido_x** y **Conocido_y**, en la ventana Argumentos de función y tras pulsar aceptar, aparecerá el resultado de la pendiente en la celda seleccionada al respecto. Igual procedimiento se utiliza para calcular la ordenada en el origen, mediante la función `INTERSECCION.EJE` y los coeficientes de correlación, para los que se utilizan las funciones `COEF.DE.CORREL` y `COEFICIENTE.R2`, según cuál sea el deseado.

Empleo de la herramienta “Regresión”

Para la mayoría de los casos las funciones comentadas anteriormente pueden ser suficientes. Sin embargo, es posible que en determinadas ocasiones se requieran hacer evaluaciones estadísticas de la regresión, como calcular valores residuales, desviaciones, probabilidades, etc. Para estas situaciones, existe una herramienta denominada “**Regresión**” que hace un completo ANOVA (Análisis de la Varianza) de la regresión lineal de los datos. Esta herramienta se encuentra en un programa de complementos de Excel que está disponible al instalar Microsoft Office o Excel y que se denomina **Análisis de datos**. Sin embargo, para usarlo en Excel es necesario cargarlo primero. Para comprobar si está cargado, hay que abrir la ficha Datos y buscar el grupo Análisis. Si este grupo no aparece, o si existiendo no muestra un comando de nombre Análisis de datos, se deberá proceder a cargar el programa de complementos **Herramientas para análisis**. Para cargar el complemento en la versión 2010 y posteriores, hay que abrir la ventana Archivo, que se muestra en el lado izquierdo de la Cinta de Opciones, en cuyo menú lateral se activará la opción Opciones que permite la apertura de la ventana de diálogo de Opciones de Excel. Una vez disponible el complemento Análisis de datos, se puede pulsar sobre él con lo que aparecerá una nueva ventana con todas las herramientas disponibles de análisis de datos. Hay que buscar en la lista la herramienta “**Regresión**” y seleccionarla para, seguidamente, aceptar pulsando el botón correspondiente. Aparecerá a continuación una ventana de diálogo donde se pide el rango de los valores de la y, el de los valores de la x y el lugar donde se quiere mostrar el informe ANOVA.

El resultado final, una vez suministrada esta información, es como el de la figura siguiente.

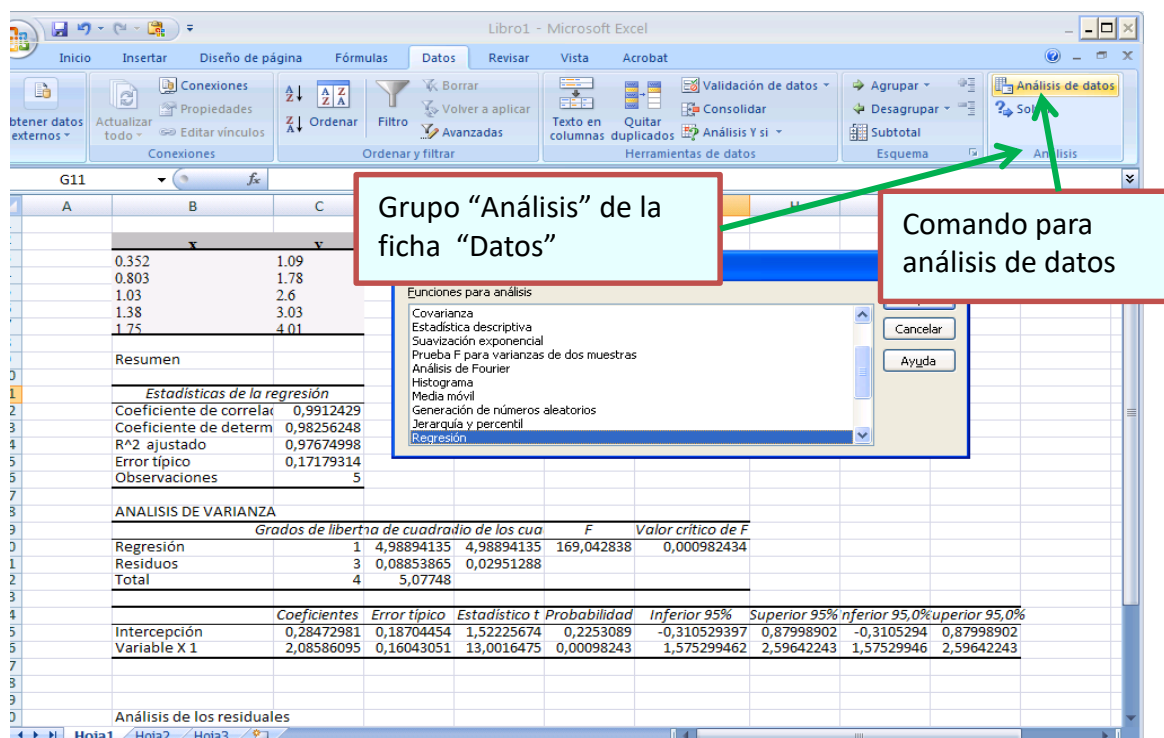


Figura 2. Resultados obtenidos tras utilizar la herramienta "Regresión" del complemento "Análisis de datos"

No obstante, si no se quiere hacer el análisis de regresión completo, se puede utilizar la función `ESTIMACION.LINEAL`, que calcula los principales parámetros de ajuste lineal, como pendiente, ordenada, sus respectivas desviaciones estándar, R^2 , entre otros. Esta función se utiliza como fórmula de matriz, por lo que para su empleo requiere seleccionar una matriz de celdas, en este caso, diez celdas en un rango de cinco filas y dos columnas. Una vez hecha la selección del rango de celdas, se introduce la función en la Barra de fórmulas y los correspondientes argumentos; dos obligatorios (*Conocido_y*, *Conocido_x*) y otros dos opcionales de carácter lógico (*Constante*, *Estadística*) que deberán tomar el valor VERDADERO en ambos casos. Una vez introducidos todos los argumentos, se presiona la combinación de teclas **CTRL+MAYÚS+ENTRAR**.

Independientemente del procedimiento utilizado para el ajuste lineal, en muchos casos además de disponer de los parámetros de ajuste, interesa tener la gráfica ajustada a los puntos. En estos casos, lo primero que hay que hacer, si no se hubiera hecho ya, es representar los puntos en un gráfico tipo XY (Dispersión). Una vez que se dispone del gráfico con los puntos, se pulsa con el botón derecho del ratón sobre alguno de los puntos representados, con lo que se desplegará un menú contextual (figura 3). Hay que seleccionar la opción Agregar línea de tendencia ..., que permite acceder a la ventana de diálogo Formato de línea de tendencia y escoger el modelo, en este caso Lineal y las opciones deseadas, como que muestre la ecuación o el coeficiente R^2 , pero teniendo cuidado con la opción Señalar intercepción =, pues esta opción forzará la línea recta a pasar por el valor que se le indique (0 por defecto). Si se ha elegido el tipo Lineal y se han marcado las casillas Presentar la ecuación en el gráfico y Presentar el valor de R cuadrado en el gráfico, el resultado que se obtendría sería como el que se muestra en segundo plano, en la figura 3.

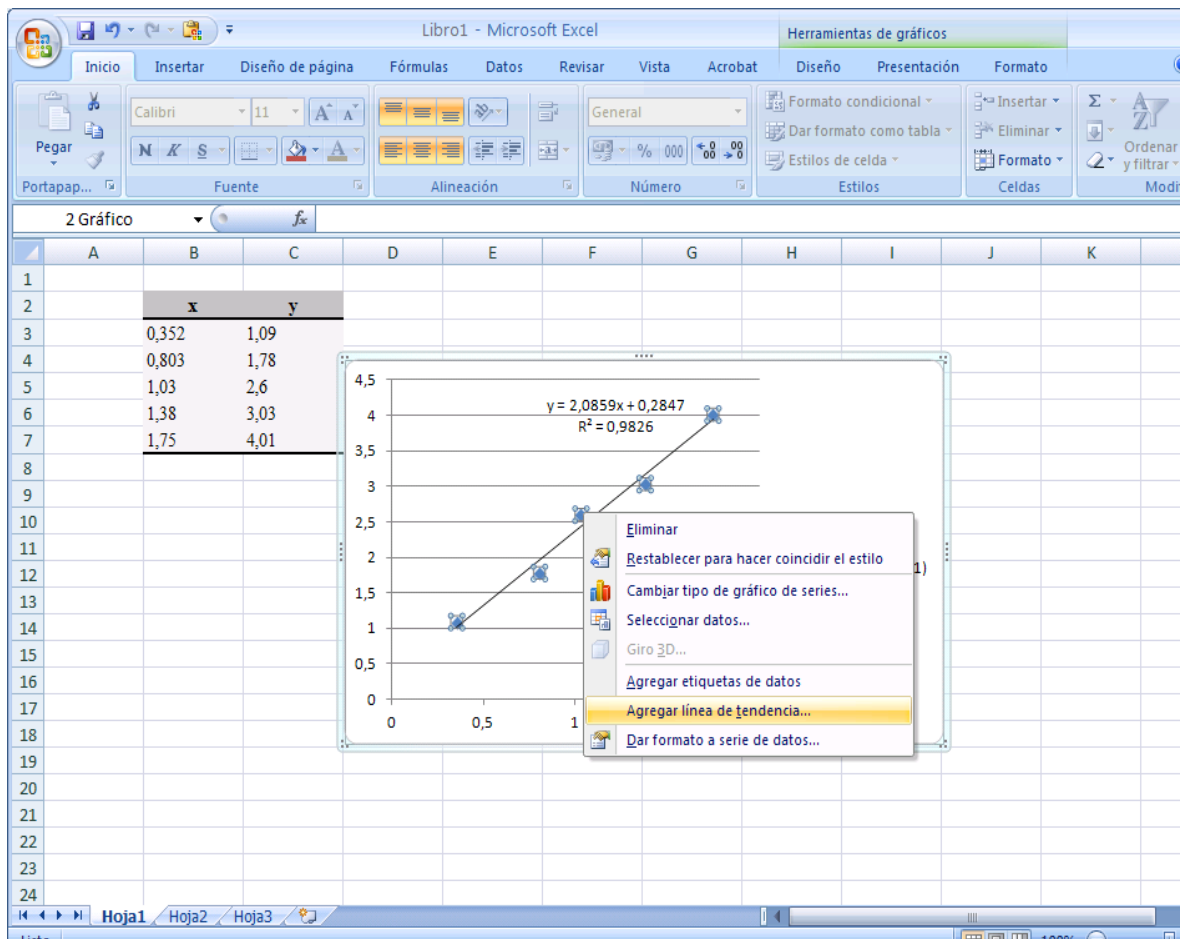


Figura 3. Representación de línea ajustada a los puntos, con la ecuación de la recta y el coeficiente de correlación al cuadrado.

Ajustes a otros modelos diferentes a la recta

Excel también ofrece la posibilidad de ajustar series de datos a otras funciones. El procedimiento para realizar estos ajustes es similar al llevado a cabo en el ejemplo de la regresión a una recta, simplemente hay que indicar el tipo de ajuste deseado (polinómico, logarítmico, etc.) entre las diferentes opciones que Excel ofrece al desplegarse la ventana Formato de línea de tendencia (Figura 4).

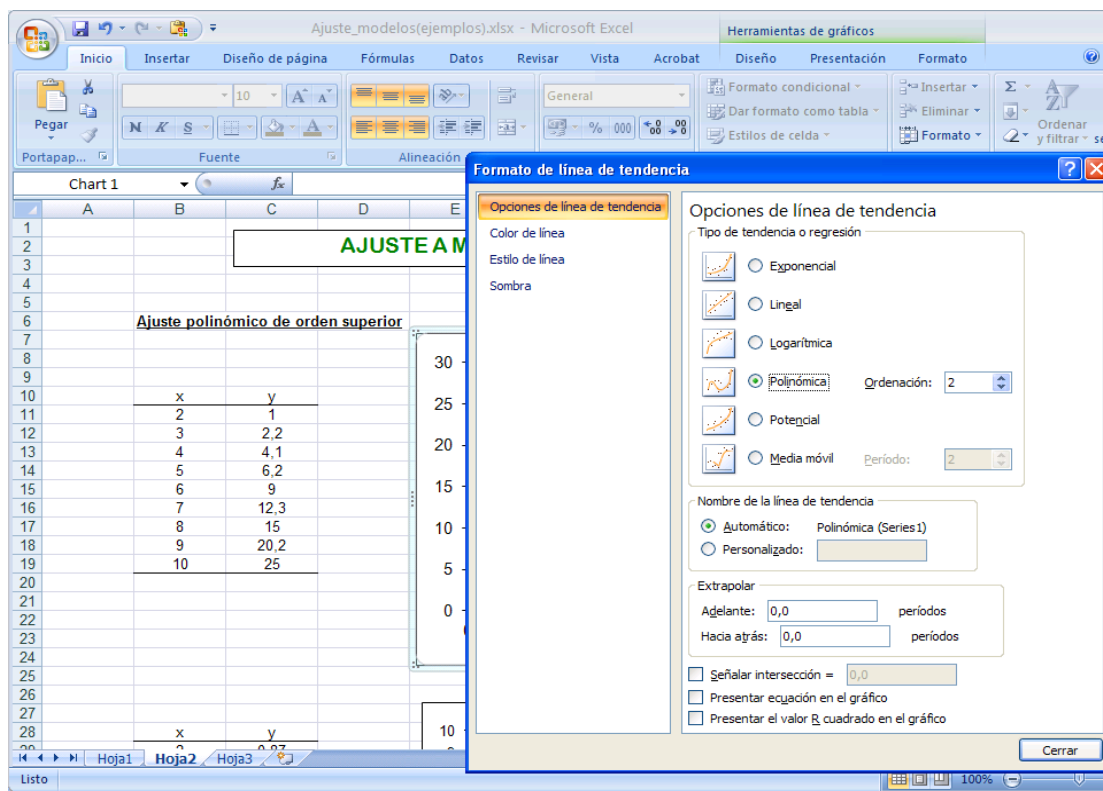


Figura 4. Posibles opciones de ajuste de datos ofrecidos por Excel.

Si en las casillas de verificación de la parte inferior de la ventana se ha seleccionado Presentar la ecuación en el gráfico y Presentar el valor de R cuadrado en el gráfico, se obtiene, junto con el gráfico, la ecuación ajustada, que puede servir para cálculos posteriores, y el coeficiente de correlación.

Ejercicios prácticos

A continuación, en este apartado, se presenta una serie de ejercicios cuya principal finalidad es poner en práctica los diferentes conceptos manejados a lo largo del tutorial y adquirir mayor soltura en el manejo de algunas de las funciones más frecuentemente utilizadas en el cálculo. Además, se introducirá al lector en el manejo de otras funciones de uso menos frecuente y en el empleo práctico de algunas de las herramientas para el análisis de datos numéricos como de optimización de funciones, enfocándolas siempre a la utilidad que podrían tener como herramientas de cálculo en química, tanto teórica como práctica. Aunque los ejemplos que vienen a continuación están explicados paso a paso y con su solución, que por supuesto no es la única, pues existen diferentes caminos para alcanzar la misma meta, sería muy recomendable que antes de ver la solución, se intente resolver el ejercicio, o al menos pensar o buscar aquellas funciones o herramientas que harían falta para conseguir llegar al objetivo.

Obtención manual de los parámetros de ajuste a una línea recta.

Como primer ejemplo práctico, se va a realizar la regresión lineal de una serie de valores o medidas realizadas durante un experimento de emisión luminiscente. La intensidad de la emisión fluorescente de las sustancias químicas que presentan esta propiedad, es directamente proporcional a la concentración de dicha sustancia, siempre dentro de unos límites. El supuesto teórico es el siguiente:

EJERCICIO 1. Se han examinado una serie de disoluciones patrón de fluoresceína en un espectrómetro de fluorescencia y han conducido a las siguientes intensidades de emisión fluorescente (en unidades arbitrarias):

Intensidad	2,0	5,0	9,1	12,5	17,2	21,0	24,8
Conc., µg/L	0,0	2,0	4,0	6,0	8,0	10,0	12,0

Determinése los parámetros de ajuste de la recta (ordenada, pendiente y coeficientes de regresión y de correlación (r y R^2)).

Como ya se ha comentado en este tutorial, Excel permite hacer todo tipo de operaciones para cálculos matemáticos, utilizando para ello operadores y funciones con las que se crean fórmulas matemáticas. En el apartado correspondiente ya se hizo un ajuste lineal mediante el empleo de las funciones estadísticas: PENDIENTE, INTERSECCION.EJE, COEF.DE.CORREL y COEFICIENTE.R2. En este otro ejemplo se verá como obtener esos mismos parámetros de ajuste de forma manual. Las ecuaciones para la realización de estos cálculos son²:

$$a = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$
$$b = \bar{y} - a\bar{x}$$
$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$
$$R^2 = r^2$$

El procedimiento manual es muy simple; tras abrir una hoja Excel se prepara una tabla como la siguiente, donde se introducen los datos y se realizan los cálculos necesarios:

² Estas ecuaciones son las que utiliza Excel para operar con las funciones PENDIENTE, INTERSECCION.EJE y COEF.DE.CORREL. Es posible encontrar otras similares en las que el parámetro b (ordenada) se obtiene de forma independiente, no a partir de la pendiente calculada, (a). Esas otras formas son equivalentes a las aquí presentadas.

Tabla 5. Tabla para la introducción de datos en el cálculo manual de los parámetros de ajuste a una recta

	x	y	$x_i - x_{med}$	$(x_i - x_{med})^2$	$y_i - y_{med}$	$(y_i - y_{med})^2$	$(x_i - x_{med})(y_i - y_{med})$
	0,0	2,0					
	2,0	5,0					
	4,0	9,1					
	6,0	12,5					
	8,0	17,2					
	10,0	21,0					
	12,0	24,8					
Sumas →							
Medias →							

Se pueden utilizar las funciones Excel SUMA y PROMEDIO para calcular la suma de los valores de x y de y , así como sus valores medios. También se puede utilizar la función SUMA para calcular los sumatorios de $x_i - x_{med}$, $y_i - y_{med}$, sus cuadrados y el producto $(x_i - x_{med})(y_i - y_{med})$. También es posible utilizar el botón **Autosuma**, representado por el signo Σ y que se encuentra en la parte izquierda de la Cinta de Opciones cuando se encuentra abierta la ficha Fórmulas, dentro del grupo Biblioteca de funciones o en la ficha Inicio en el grupo Modificar. Activando la celda donde se quiere colocar el resultado y pulsando con el ratón directamente sobre el botón, se inserta la función SUMA a la vez que Excel propone el rango de celdas que se utilizarán para el cálculo. En caso de estar conforme, simplemente se acepta pulsando la tecla **Intro**, [↵], mientras que, en caso contrario, se puede modificar el rango de celdas. Junto al botón **Autosuma** se encuentra un comando de lista desplegable (punta de flecha en forma de triángulo con el vértice hacia abajo) donde se pueden seleccionar otras funciones de uso frecuente, como PROMEDIO, CONTAR NUMEROS, MAX y MIN, muy útiles en numerosas operaciones de estadística descriptiva. Una vez que se ha completado la tabla utilizando las funciones y fórmulas adecuadas a cada caso, se está en disposición de realizar los cálculos de la pendiente, la ordenada y el coeficiente de regresión y de correlación.

Ajuste lineal mediante regresión matricial

En el presente epígrafe se va a abordar otra forma de calcular los parámetros de ajuste de una ecuación dada, a la vez que permitirá introducirse en el empleo de Excel para realizar operaciones y cálculos con matrices. Cada uno de los pares de valores (x,y) del ejercicio del epígrafe anterior representa la relación entre una variable independiente, x , en este caso la concentración de fluoresceína, y la variable dependiente, y , la fluorescencia medida para esa concentración de fluoresceína. Como se parte del supuesto teórico de que la relación entre x e y es lineal, ambas variables se relacionarán mediante la ecuación:

$$y = ax + b$$

de forma que los siete pares de valores pueden relacionarse mediante las siguientes siete ecuaciones, que en su conjunto conforman un sistema lineal de ecuaciones:

$$\begin{aligned}
2,0 &= a \cdot 0,0 + b \\
5,0 &= a \cdot 2,0 + b \\
9,1 &= a \cdot 4,0 + b \\
12,5 &= a \cdot 6,0 + b \\
&\dots\dots\dots \\
24,8 &= a \cdot 12,0 + b
\end{aligned}$$

donde a y b , como ya se ha comentado, son los coeficientes de ajuste que es preciso calcular para establecer el modelo lineal que posteriormente permitirá, mediante interpolación o extrapolación, estimar otros datos.

Este sistema lineal de ecuaciones también se puede expresar en forma de ecuación matricial, que para este caso particular toma la forma:

$$\begin{pmatrix} 2,0 \\ 5,0 \\ 9,1 \\ 12,5 \\ \vdots \\ 24,8 \end{pmatrix} = \begin{pmatrix} 0,0 & 1 \\ 2,0 & 1 \\ 4,0 & 1 \\ 6,0 & 1 \\ \vdots & \vdots \\ 12,0 & 1 \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix}$$

siendo la forma general, para todos los casos similares a este:

$$\mathbf{y} = \mathbf{X} \cdot \mathbf{c}$$

donde \mathbf{y} (en minúscula y negrita) representa el vector (matriz de una sola columna) que contiene las variables dependientes; \mathbf{X} (en mayúscula y negrita) la matriz $n \times 2$ que contiene una columna con las variables independientes y otra formada por “unos”, tantos como variables independientes se hayan establecido. Finalmente, el vector matricial \mathbf{c} contiene los valores de los coeficientes a y b que hay que calcular. Puesto que el vector matricial \mathbf{c} contiene los coeficientes, es preciso despejarlo de la ecuación matricial para poder determinar los valores numéricos de los elementos que contiene (a y b). El procedimiento matemático para llevarlo a cabo se denomina **regresión matricial** y consta de los siguientes pasos que a continuación se comentan brevemente:

Paso 1. Se multiplica por la transpuesta de \mathbf{X} (\mathbf{X}^T) ambos términos de la igualdad. La multiplicación se hará por la izquierda para mantener la congruencia de las dimensiones de las matrices.

$$\mathbf{X}^T \cdot \mathbf{y} = \mathbf{X}^T \cdot \mathbf{X} \cdot \mathbf{c}$$

Paso 2. Se obtiene la matriz inversa de la matriz resultante de multiplicar $\mathbf{X}^T \cdot \mathbf{X}$ y se multiplica por dicha matriz ambos términos de la igualdad. Como en el paso anterior la multiplicación debe de hacerse por la izquierda

$$(\mathbf{X}^T \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^T \cdot \mathbf{y} = (\mathbf{X}^T \cdot \mathbf{X})^{-1} \cdot (\mathbf{X}^T \cdot \mathbf{X}) \cdot \mathbf{c}$$

El producto de una matriz por su inversa (en este caso $(X^T \cdot X)^{-1} \cdot (X^T \cdot X)$) genera como resultado la matriz identidad, I , que, en este caso particular, sería:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

y que tiene la propiedad de ser el elemento neutro del producto de matrices. Esto quiere decir que el producto de cualquier matriz por la matriz identidad no tiene ningún efecto con lo que al final, la ecuación matricial del paso 2 se puede escribir como:

$$c = (X^T \cdot X)^{-1} \cdot X^T \cdot y$$

Excel dispone de varias funciones para operar con matrices, obtener sus transpuestas o inversas, calcular su valor numérico (determinante de la matriz), etc. que permitirán calcular los valores de los parámetros de ajuste del ejercicio de ajuste lineal del epígrafe anterior. En primer lugar, es necesario introducir los datos en una hoja Excel y definir las matrices X e y , como se muestra en la figura.

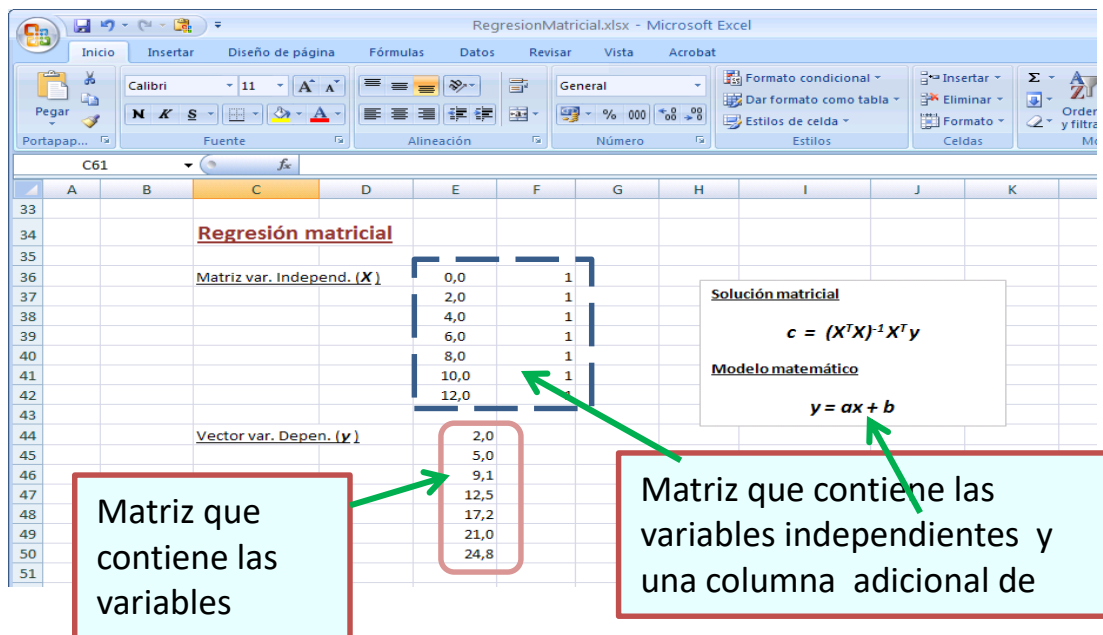


Figura 5. Hoja Excel con los datos necesarios para el ajuste lineal. Incluye un cuadro de texto con información sobre el modelo matemático al que se va a hacer el ajuste y la ecuación matricial que generará los coeficientes a y b .

Mientras que en el cálculo manual de los coeficientes a y b no es necesario definir previamente el modelo matemático, aquí es muy importante, pues en el caso que se hubiera establecido como modelo lineal $y = a + bx$, la columna de “unos” de la matriz de variables independientes debería haberse situado a la izquierda de la que contiene los valores de x . Una vez que se han introducido los datos en la hoja Excel, el siguiente paso es obtener la matriz transpuesta X^T . Se puede hacer de forma manual, introduciendo en un grupo de celdas los datos de la primera columna, pero ahora en una fila y los de la segunda columna en la

fila inmediatamente inferior; sin embargo, este procedimiento manual aparte de ser laborioso, no está exento de posibles errores que lleven a introducir algún dato de forma equivocada. Excel dispone de la función **TRANSPONER** que realiza esta operación de forma automática y que utiliza como argumento el rango de celdas que contiene la matriz original. Para obtener la transpuesta se seleccionará una celda de la hoja y se introducirá la función con el argumento. A partir de esta celda se seleccionará un rango de celdas de iguales dimensiones que la matriz resultante, es decir para este ejemplo, se seleccionará un grupo de celdas de dimensiones 2 x 7. Una vez seleccionadas es necesario presionar la combinación de teclas **CTRL+MAYÚS+ENTRAR**. De no hacerse así, Excel solo mostrará el primer elemento de la matriz transpuesta. Si se opera correctamente, el resultado deberá ser el de la figura 6

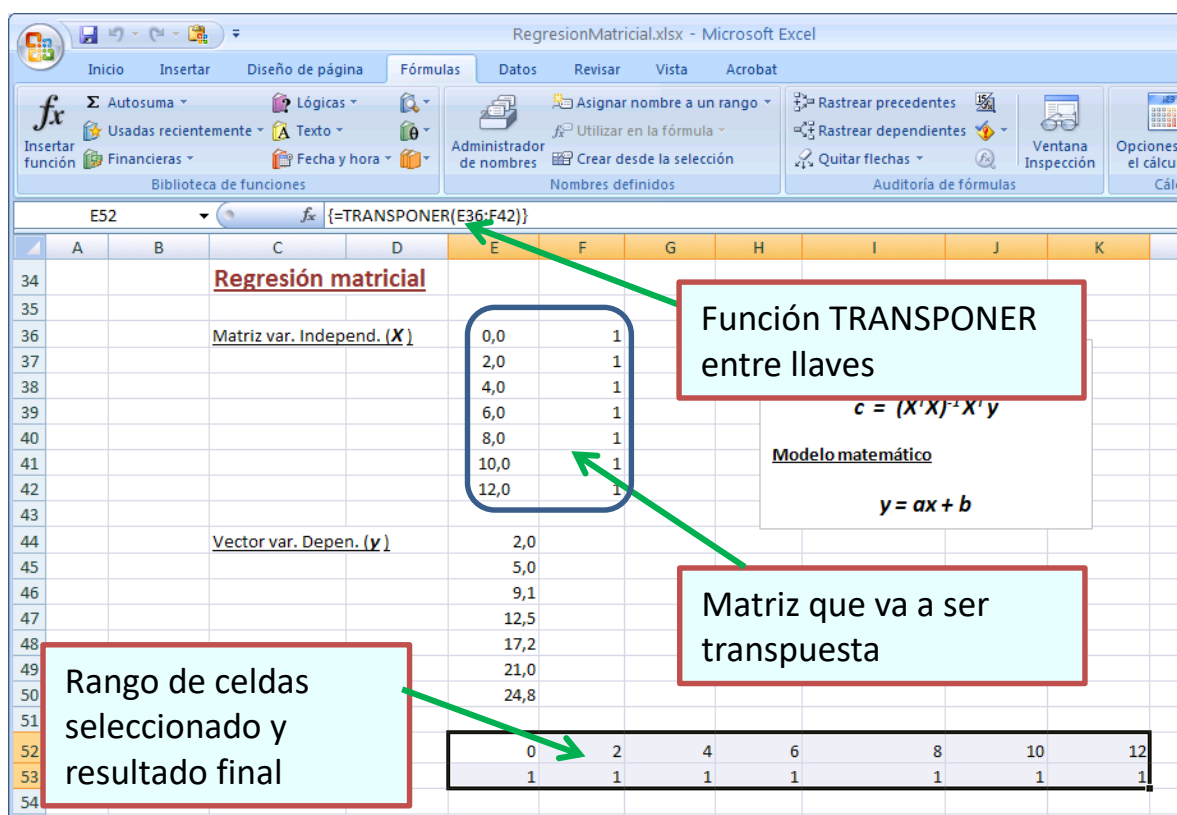


Figura 6. Hoja Excel mostrando la matriz a transponer y el resultado final que se obtiene tras aplicar la función **TRANSPONER**

La función **TRANSPONER** es lo que se denomina una función de matriz. En su mayor parte, las fórmulas de matriz usan sintaxis de fórmula estándar, comienzan con un signo igual y se puede utilizar cualquiera de las funciones incorporadas de Excel. La principal diferencia es que al utilizar una fórmula de matriz es necesario presionar **CTRL+MAYÚS+ENTRAR** para especificarla. Al hacer esto, Excel incluye la fórmula de matriz entre llaves. Esta operación solo la puede hacer Excel, de forma que, si se escribe las llaves manualmente, la fórmula se convertiría en una cadena de texto y no funcionaría.

Por supuesto, además de la función `TRANSPONER` existen otras funciones de matriz que permiten realizar otros cálculos. El siguiente paso en este ejercicio de regresión matricial sería obtener el producto $X^T \cdot X$, por lo que será necesario multiplicar la matriz transpuesta obtenida anteriormente por la matriz que contenía las variables independientes. Para el producto de matrices existe la función `MMULT` que tiene los argumentos **Matriz1** y **Matriz2**. **Matriz1** es el rango de celdas que contiene, en este caso la matriz X^T y **Matriz2** el correspondiente a la matriz X . De acuerdo con las reglas del producto de matrices, la matriz resultante de multiplicar una matriz de dimensiones 2 x 7 por otra de 7 x 2 deberá tener dimensiones 2 x 2, por lo que se seleccionará un rango de celdas de estas dimensiones, antes de pulsar la combinación `CTRL+MAYÚS+ENTRAR`. Similar procedimiento se deberá seguir para continuar el proceso de regresión matricial. Para obtener la matriz inversa de $X^T \cdot X$ se utiliza la función `MINVERSA`. El valor del coeficiente de correlación, R^2 . Se puede calcular utilizando la función estadística `COEFICIENTE.R2` o a partir de la ecuación:

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}$$

donde \hat{y}_i representa cada uno de los valores y , estimados mediante el modelo matemático obtenido; y_i los valores originales de la variable dependiente, e \bar{y} el valor medio de estos. Los valores estimados, \hat{y}_i se pueden obtener uno a uno sustituyendo en la ecuación ajustada los diferentes valores de x , o de forma mucho más rápida, de una sola vez, multiplicando la matriz X por la matriz vectorial c .

Como puede verse en estos ejemplos de regresión, Excel permite llegar a los mismos resultados a través de diferentes caminos, unos más directos que otros. Es obvio que en el caso de tener que calcular los parámetros de ajuste lo más cómodo sería utilizar las funciones de estimación directa de los parámetros de ajuste. Sin embargo, existen situaciones, como las del siguiente ejemplo, en que ello no es posible.

***EJERCICIO 2.** En una serie de experimentos se ha encontrado que las medidas realizadas son función de dos variables, x_1 y x_2 que afectan simultáneamente a los resultados de acuerdo a un modelo lineal del tipo:*

$$y = a + bx_1 + cx_2 + dx_1x_2$$

Se han llevado a cabo diferentes experimentos para diferentes valores de x_1 y x_2 obteniéndose los siguientes resultados:

Resultado	3,4	5,1	6,6	4,9	5,1	6,1	4,4	4,9	5,4
x_1	1	1	1	2	2	2	3	3	3
x_2	1	2	3	1	2	3	1	2	3

*Determinese los coeficientes de ajuste, **a**, **b**, **c**, **d** y el coeficiente de correlación, R^2 .*

Como puede verse, se trata de una función lineal de dos variables con interacciones entre ellas. Sin embargo, su solución matricial es exactamente igual al del ejemplo anterior de una sola variable:

$$c = (X^T \cdot X)^{-1} \cdot X^T \cdot y$$

solo que ahora el vector matricial \mathbf{c} estará formado por cuatro elementos, uno por cada coeficiente a , b , c , d . También es diferente la matriz \mathbf{X} que ahora tendrá cuatro columnas, una por cada coeficiente de ajuste, tal y como se muestra en la figura.

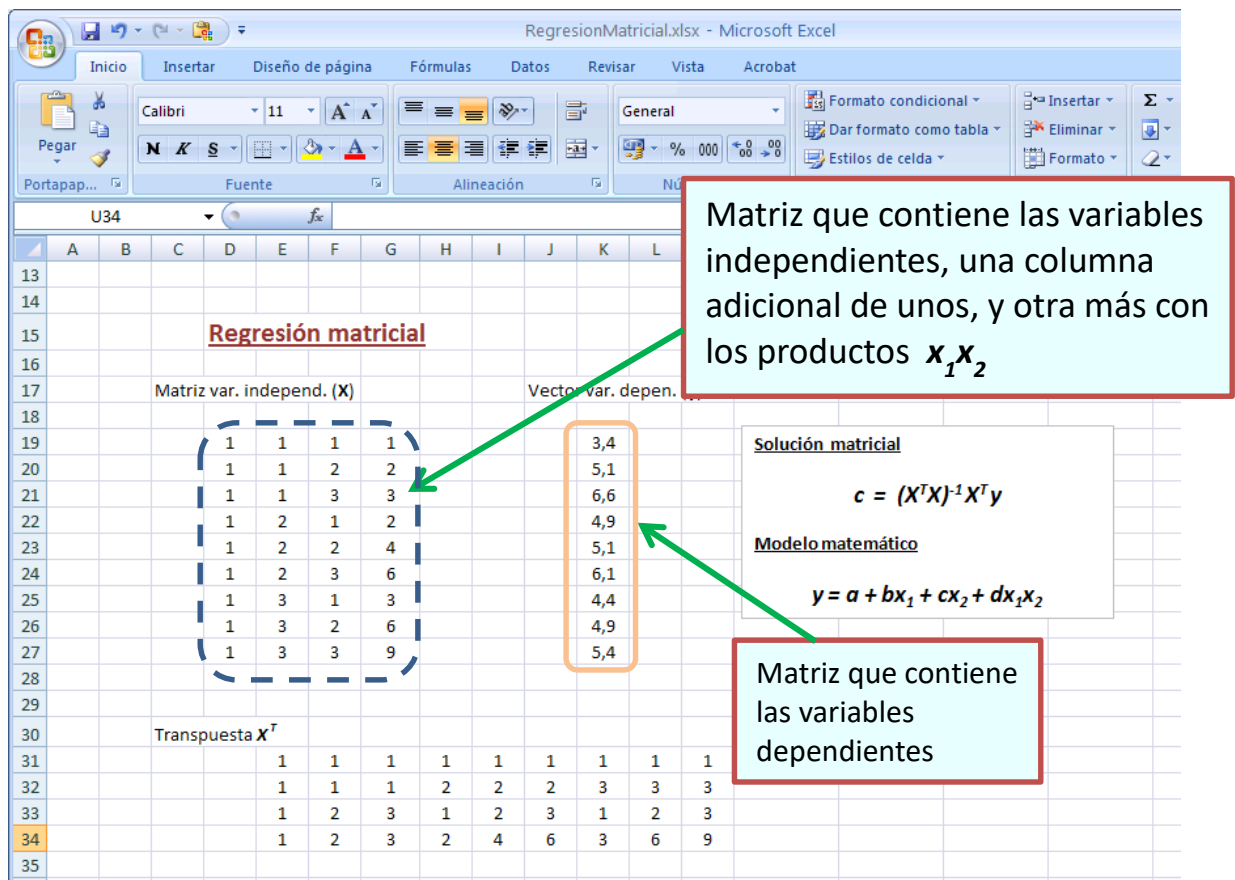


Figura 7. Hoja Excel con los datos necesarios para el ajuste lineal del ejemplo.

En este caso la primera columna es la que contiene los “unos” que representan a la variable independiente a , ya que en el modelo matemático general al que se van a ajustar los datos, este coeficiente es el primer término de dicha ecuación. Puesto que el coeficiente b es el término que multiplica a la variable x_1 , y se encuentra en segundo lugar de la ecuación, la segunda columna de la matriz \mathbf{X} deberá contener los valores de la variable independiente x_1 . El mismo criterio se sigue para la columna tercera y cuarta. Esta última, como representa el producto $x_1 \cdot x_2$, estará formada por los productos de los valores de dichas variables. El resto del proceso para obtener los coeficientes de regresión es exactamente igual que en el ejemplo univariante.

Resolución de ecuaciones mediante métodos iterativos

Excel dispone de una herramienta denominada **Buscar objetivo**, que mediante procedimientos iterativos es capaz de buscar una solución a la ecuación con un grado de aproximación aceptable, generalmente definido por el usuario. Para practicar y entender cómo funciona esta herramienta se propone el siguiente ejercicio relacionado con la emisión fluorescente, que como ya se ha comentado, solo es directamente proporcional a la

concentración de la sustancia emisora a concentraciones muy bajas. En realidad, la intensidad de emisión de fluorescencia molecular se ajusta a la función exponencial tipo:

$$I_f = k \cdot \phi_f \cdot I_0(1 - e^{-a \cdot l \cdot C})$$

donde k es una constante de proporcionalidad que depende del instrumento de medida; ϕ_f el rendimiento cuántico de la fluorescencia, a es la absorptividad de la molécula y l el paso óptico de la célula de medida.

EJERCICIO 3. La intensidad relativa de emisión fluorescente de la fluoresceína a concentraciones superiores a las analíticas se ajusta a la expresión:

$$I_f = 622 \cdot (1 - e^{-0,00351C}).$$

Calcúlese la concentración, C que tiene que tener una disolución de fluoresceína para que produzca una emisión de intensidad igual a 150 unidades.

Como se puede apreciar, no es fácil despejar la variable C de la ecuación que la relaciona con la intensidad de emisión. Para calcularla, de forma aproximada, se puede utilizar, como ya se ha comentado la herramienta Buscar objetivo. Como paso previo es necesario configurar la hoja de cálculo, definiendo la celda que contendrá la fórmula matemática y la que contendrá la concentración (figura 8). Una vez que se ha configurado la hoja se activará la herramienta Buscar objetivo, que se encuentra en el grupo Herramientas de datos de la ficha Datos. Para utilizar esta herramienta es necesario desplegar el menú Análisis Y si, y seleccionar la herramienta. Se abre una pequeña ventana de diálogo solicitando la celda que contiene la fórmula matemática, el valor que tiene que tomar dicha función y la celda donde se colocará el resultado (figura 8). Una vez que se han introducido los datos requerido, tras presionar **Aceptar**, la ventana flotante se sustituye por otra de similar tamaño con un mensaje que indica si se ha encontrado una solución, así como el valor buscado y el encontrado, que pueden llegar a ser diferentes, en función de la precisión de búsqueda de objetivo que vengan determinadas en Opciones de Excel. Por defecto, Excel realiza un máximo de 100 iteraciones para alcanzar el objetivo con una precisión de 0,001. El número máximo de iteraciones a realizar, así como la precisión deseada, pueden ser modificadas por el usuario. Recuérdese que al menú Opciones de Excel se accede desde la ventana Archivo, en la versión 2010 y posteriores. Una vez abierta la ventana Opciones de Excel, los valores de precisión y número de iteraciones se encuentran en el menú Fórmulas. También es necesario activar la casilla Habilitar cálculo iterativo.

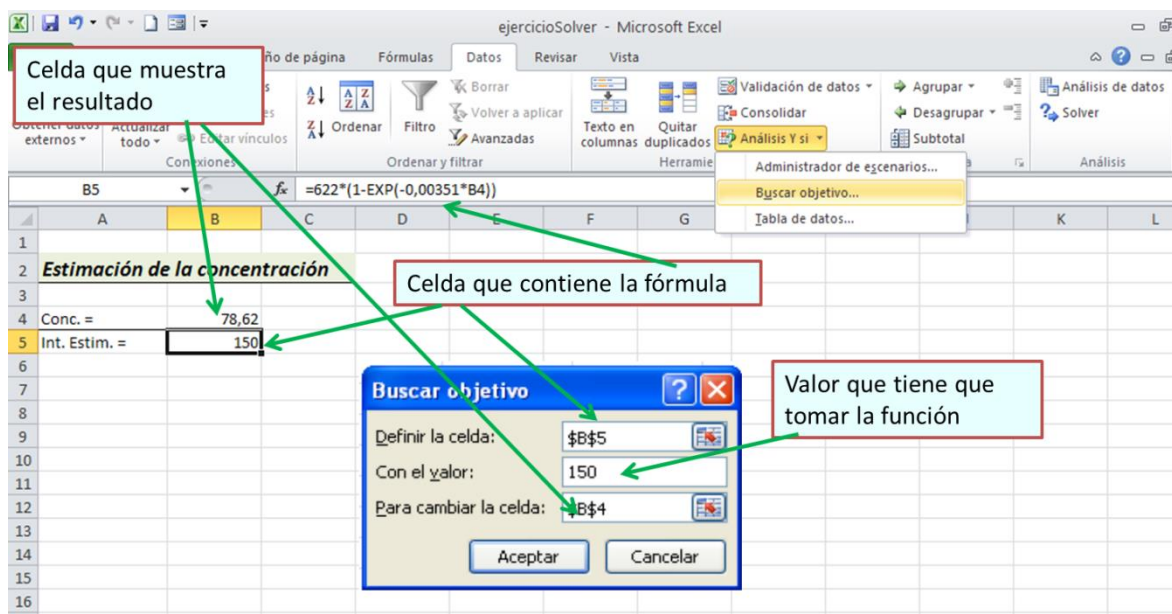


Figura 8. Hoja Excel implementada para el cálculo de la concentración a partir de la ecuación de la intensidad de emisión fluorescente. Muestra la ventana de diálogo *Buscar objetivo*.

Problemas y ejercicios para resolver con Excel

Adicionalmente a los ejercicios anteriores, cuya resolución se muestra paso a paso a la vez que se comenta los fundamentos teóricos y las estrategias seguidas para poder ser resueltos, se presenta a continuación algunos ejercicios y supuestos prácticos que pueden ser resueltos mediante el empleo de funciones y herramientas de **Excel**, siguiendo las orientaciones mostradas anteriormente.

1 En una práctica de determinación de potasio en vino mediante emisión de llama, un grupo de alumnos realizó un calibrado con los datos de la tabla, y siguiendo las directrices del guion de prácticas lo ajustaron a una línea recta. Compruébese si la elección es correcta y en caso contrario ajústese los datos experimentales a un modelo más adecuado.

Conc. mg/L	1,0	2,0	4,0	6,0	8,0	10,0	12,0	14,0	18,0
I. emis.	0,114	0,186	0,315	0,403	0,492	0,564	0,629	0,694	0,780

(NOTA. Se considera que un modelo es adecuado cuando los datos ajustan a dicho modelo en más del 99%)

2 Se sabe que la conductividad de las disoluciones acuosas de hidróxido sódico se relaciona de forma lineal con la concentración de esta base en el intervalo comprendido entre 0,1 y 0,0001 M. La siguiente tabla muestra las medidas realizadas a una serie de disoluciones patrón de hidróxido sódico.

NaOH (M)	0,03	0,02	0,015	0,01	0,005	0,002
Conduc. ($\mu\text{S}\cdot\text{cm}^{-1}$)	6530	4320	3390	2272	1423	650

Represente los valores y decida a qué modelo matemático se ajustan. Obtenga la función matemática que relaciona conductividad con concentración de NaOH. Si la conductividad de una disolución de NaOH presenta una conductividad de $3953 (\mu\text{S}\cdot\text{cm}^{-1})$, cuál es su concentración. ¿Qué conductividad se espera de una disolución de NaOH 0,012 M?

3 En un estudio de validación de un método bioanalítico se adicionaron diferentes cantidades de analito A sobre plasma sanguíneo y seguidamente se procedió a analizar las muestras, obteniéndose los siguientes resultados:

Concentrac. A	1	4	40	80	160
Resp. instrum.	12	32	242	486	882

Compruébese si la respuesta instrumental se encuentra relacionada linealmente con la concentración. En caso contrario indíquese cuál sería el tipo de función más adecuada.

4 En un experimento de química se está estudiando la relación existente entre la cantidad de materia transferida desde un cromatógrafo a un espectrómetro de masas y la señal proporcionada por este. En la siguiente tabla se muestran los resultados obtenidos:

Cantidad (ng)	2200	1760	1430	1100	660	440	220	110	44
Señal	9989	7685	5723	3860	2132	1391	710	290	201

Represente los resultados y deduzca a qué modelo matemático se ajustan. Obtenga la ecuación de la función matemática y deduzca cuál sería la señal medida en el aparato si se introdujeran 320 ng de compuesto. (NOTA. Considérese el comentario del ejercicio RMM-1 para decidir cuándo un modelo ajusta suficientemente a los datos experimentales)

5 Se introduce 1 mol de Ar gaseoso en un cilindro con un pistón. Cuando el pistón comprime el gas a temperatura constante, la presión dentro del recipiente se modifica de acuerdo a la siguiente tabla:

P (mm Hg)	100	200	300	400	600
V (L)	250	130	85	64	60

Suponiendo un comportamiento ideal del gas, calcule a que temperatura se han llevado a cabo estas medidas. Dato: $R = 0,082 \text{ atm}\cdot\text{L}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$.

6 Se sabe que la absorbancia (A), de radiación a una cierta longitud de onda es proporcional a la concentración, c, y están relacionadas mediante la expresión:

$$A = \varepsilon l c$$

Los datos que se dan a continuación corresponden a medidas de absorbancia a diferentes concentraciones de un complejo coloreado MLn.

$10^4 \cdot [\text{ML}_n] \text{ (M)}$	0,5	1,5	3,0	6,0	9,0	12,0
Absorbancia	0,034	0,131	0,221	0,441	0,697	0,890

Considerando que la longitud de la célula de medida del colorímetro es de 2 cm, ($l = 2$) determine el valor de ε con sus límites de confianza. De acuerdo con el modelo matemático del enunciado, la ordenada en el origen debería ser nula. ¿Podría decirse que los resultados obtenidos confirman esa hipótesis? ($\alpha = 0,05$)

7 La ecuación de Antoine es una ecuación empírica que proporciona la presión de vapor (en mm Hg) de un compuesto en función de tres parámetros y de la temperatura en grados centígrados, pero no informa acerca del valor de la entalpía de vaporización. Sabiendo que dicha ecuación toma la expresión:

$$\log_{10}(P) = A - \frac{B}{C + t}$$

y que los parámetros de la ecuación de Antoine para el agua son:

	A	B	C	Intervalo de validez ($^{\circ}\text{C}$)
Agua	8.07131	1730.63	233.426	2 - 100

prepare una hoja de cálculo Excel que permita obtener los datos de presión de vapor del agua en función de la temperatura en el intervalo de validez especificado (calcule la presión de vapor cada dos grados centígrados). Sabiendo que la presión de vapor (en mm de Hg) se relaciona con la temperatura absoluta a través de la ecuación:

$$\ln P = -\frac{\Delta H_v}{R} \frac{1}{T} + C$$

Obténgase la entalpía de vaporización (ΔH_v) para el agua. (Dato. $R = 8.31 \text{ J mol}^{-1} \text{ K}^{-1}$).

8 Los siguientes datos corresponden a las medidas de la temperatura de ebullición del alcohol etílico en función de la presión del sistema.

T, °C	20	30	40	50	60	70	74
p, mmHg	50	90	146	186	237	330	370

Utilizando la ecuación y dato del ejercicio anterior, determínese la variación de la entalpía estándar de vaporización (ΔH_v) con sus límites de confianza. Busque el valor en alguna base de datos, por ejemplo en la base NIST (<http://webbook.nist.gov/chemistry/index.html>) y determine si las diferencias entre el valor tabulado y el experimental es significativo al 95% ($\alpha = 0.05$)

9 Se ha medido la temperatura de ebullición del agua en un sistema sometido a vacío parcial, obteniéndose las temperaturas que se indican en la tabla:

T, °C	36	47	60	65	70	78	81	86	89	93
p, torr	50	90	146	186	237	330	370	450	510	590

Determínese la variación de la entalpía de vaporización del agua estándar (ΔH_v°) con sus límites de confianza. ¿Existen diferencias significativas con respecto al valor tabulado (40,8 kJ/mol)?

10 En un estudio del equilibrio de solubilidad del bórax se encontraron los siguientes valores de la constante del producto de solubilidad, K_{ps} en función de la temperatura:

T (K)	294	297	303	309	313
K_{ps}	$4,050 \times 10^{-3}$	$7,160 \times 10^{-3}$	$1,560 \times 10^{-2}$	$2,978 \times 10^{-2}$	$6,537 \times 10^{-2}$

Sabiendo que la constante de equilibrio de una reacción se relaciona con la temperatura absoluta por medio de la siguiente expresión:

$$\ln K = - \frac{\Delta H^\circ}{R T} + \frac{\Delta S^\circ}{R}$$

Determínense la variación de entalpía (ΔH°) y de la entropía (ΔS°) estándar de la reacción indicada, con sus respectivos límites de confianza.
($R = 8.31 \text{ J mol}^{-1} \text{ K}^{-1}$).

11 Para la reacción $2\text{SO}_2(\text{g}) + \text{O}_2(\text{g}) \rightleftharpoons 2\text{SO}_3(\text{g})$ se ha podido medir la constante de equilibrio a diferentes temperaturas, obteniéndose los valores que figuran a continuación:

T (K)	800	850	900	950	1000	1050	1100
K_p	930	175	45	60	3,5	1	0,43

Utilizando la ecuación y dato del ejercicio anterior, determínese la variación de la entalpía estándar (ΔH°) de la reacción indicada. Exprésela con un número correcto de cifras significativas.

12 Para la reacción de descomposición $C \rightarrow A + B$ se ha determinado la constante cinética a distintas temperaturas, obteniéndose los datos de la siguiente tabla:

t (°C)	24,8	29,9	34,7	40,2
$10^3 k$ (s ⁻¹)	0,900	2,18	4,82	11,0

La constante cinética, k , y la temperatura absoluta se relacionan entre sí mediante la ecuación:

$$\ln K = \ln A - \frac{E_a}{RT}$$

Calcule la energía de activación de dicha reacción y estime el valor de la constante cinética a 32 °C. ($R = 8.31 \text{ J mol}^{-1} \text{ K}^{-1}$).

13 En un experimento de química se encontró que el resultado, y , era función de dos variables, x_1 y x_2 y que la función $f(x_1, x_2)$ que las relaciona, toma la forma general $y = a + bx_1 + cx_2 - dx_1x_2 + ex_1^2 + fx_2^2$. A partir de los datos de la tabla obténgase el modelo matemático

x_1	1	1	1	2	2	2	3	3	3
x_2	1	2	3	1	2	3	1	2	3
y	8	14	24	12	17	26	18	22	30

14 En cromatografía, la altura equivalente de plato teórico (AETP) para un determinado soluto retenido en la columna, está relacionada con la velocidad de desplazamiento de la fase móvil en el interior de la columna. La ecuación, conocida como ecuación de van Deemter toma la forma general:

$$H = A + \frac{B}{u} + Cu$$

Donde u representa la velocidad media de la fase móvil y H la AETP. En una serie de experimentos se ha calculado la AETP para las diferentes velocidades medias que se muestran en la tabla. Encuentre los valores que toman A , B y C en la ecuación que ajustan a estos datos

u , mm/s	0,62	1,24	1,86	2,48	3,10	3,73
H , mm	0,91	0,69	0,66	0,67	0,70	0,74

15 Para determinar el contenido de Fe(III) en una muestra de agua se utilizó el método de las adiciones estándar. Para ello se preparó una serie de 5 matraces volumétricos de 10,0 mL. En todos ellos se colocó un mL de muestra. A cada matraz se añadió un volumen de 0,0; 1,0; 2,0; 3,0 y 4,0 mL de una disolución patrón 10,2 mg/L de Fe(III), tras lo cual, se enrasaron los matraces con disolución $1,0 \times 10^{-2}$ M de tiocianato. La absorbancia del complejo $\text{Fe}(\text{SCN})^{2+}$ formado en cada matraz se midió a 480 nm siendo las lecturas del espectrofotómetro 0,240; 0,437; 0,621; 0,809 y 1,009, respectivamente. ¿Cuál es la concentración de Fe(III) en la muestra de agua?

16 En un experimento se obtuvieron los resultados que se indican en la tabla, los cuales se ajustaron mediante regresión lineal al polinomio de tercer orden $y = 16,1x^3 - 33,9x^2 + 29,3x - 0,61$. Utilizando las opciones de Excel para operar con matrices calcule el coeficiente de correlación múltiple, R^2 .

x	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
y	-0,45	1,98	3,28	5,16	7,08	7,34	7,66	9,56	9,48	9,30	11,2